Journal of the Operations Research
Society of Japan
Vol. 39, No. 2, June 1996

PARTIAL PROXIMAL METHOD OF MULTIPLIERS FOR CONVEX PROGRAMMING PROBLEMS

Satoru Ibaraki   Masao Fukushima
Kyoto University

(Received June 13, 1994)

Abstract   Two variants of the partial proximal method of multipliers are proposed for solving convex programming problems with linear constraints, where the objective function is expressed as the sum of two convex functions. The iteration of each algorithm consists of computing an approximate saddle point of the argumented Lagrangian. The global convergence is established under an approximation criterion for computing the saddle point. In particular, for the convex programming problem with multiple set constraints and the traffic assignment problem, one of the proposed algorithms can effectively be implemented on a parallel computer.

1. Introduction

With the development of computer science, parallel and distributed computation has been extensively studied (e.g. [2]). For convex programming problems, many researchers have proposed parallel algorithms based on the method of multipliers [13], the proximal point algorithm [4, 14, 15, 21], the splitting algorithm [7, 9, 20], the alternating direction method of multipliers [6, 8] and the modified trust region method [10]. On the other hand, Ha [12] presented a modification of the proximal point algorithm, in which only some of the variables are involved in the proximal term. This partial proximal method has been further analyzed by Bertsekas and Tseng [3], who particularly show that partial proximal minimization algorithms are closely related to some parallel algorithms in convex programming.

Let $F : R^n \to R \cup \{+\infty\}$ be a closed proper convex function, A an $m \times n$ matrix and b an m-dimensional vector. Consider the following convex programming problem:

$$\begin{array}{ll} \text{minimize} & F(z) \\ \text{subject to} & Az = b. \end{array} \tag{1.1}$$

The Lagrangian function $\ell : R^{n+m} \to R \cup \{+\infty\}$ is defined by

$$\ell(z,p) = F(z) - \langle p, Az - b\rangle, \tag{1.2}$$

where $\langle \cdot, \cdot \rangle$ denotes the inner product. It is well known [18, Theorem 28.3] that, under appropriate conditions, a saddle point (z^*, p^*) of the function ℓ is a pair of optimal solutions for problem (1.1) and its dual.

Among other things, we shall in particular be interested in the case where the objective function F is separable in two groups of variables, i.e.,

$$F(z) = F(x,y) = f(x) + g(y),$$

where $z = (x,y)$ ($x \in R^{n_1}$, $y \in R^{n_2}$, $n_1 + n_2 = n$) and the function g is strongly convex with modulus β, that is, there exists a positive constant β such that

$$g((1-\lambda)y+\lambda y') \leq (1-\lambda)g(y)+\lambda g(y') - \frac{1}{2}\beta\lambda(1-\lambda)|y-y'|^2 \tag{1.3}$$

for all y, $y' \in R^{n_2}$ and $\lambda \in (0,1)$. Then problem (1.1) is rewritten as

$$\begin{array}{ll} \text{minimize} & f(x)+g(y) \\ \text{subject to} & A_1x + A_2y = b, \end{array} \tag{1.4}$$

where $A = (A_1, A_2)$.

The purpose of this paper is to propose two variants of the partial proximal method of multipliers for problem (1.4) and prove their global convergence. In particular, for problems with separable structure, one of the proposed algorithms can be shown to incorporate the separability and effectively be implemented on a parallel computer.

The paper is organized as follows. In Section 2, we review the method of multipliers and the proximal method of multipliers for problem (1.1). In Section 3, we propose two variants of the partial proximal method of multipliers designed to solve problem (1.4). In Section 4, we give some basic results. In Section 5, we establish convergence theorems for the proposed algorithms. Moreover, we apply one of the proposed algorithms to the convex programming problem with multiple set constraints and the traffic assignment problem in Sections 6 and 7, respectively.

2. Preliminaries

A variety of methods have been developed for finding a saddle point of the Lagrangian function ℓ defined by (1.2). In particular, the method of multipliers (MOM) generates a sequence $\{(z^{(\mu)}, p^{(\mu)})\}$ converging to a saddle point of ℓ by the following iterative scheme:

$$(z^{(\mu+1)}, p^{(\mu+1)}) \approx \arg\min_{z\in R^n}\left\{\max_{p\in R^m} L_1(z,p;p^{(\mu)},\gamma^{(\mu)})\right\}, \tag{2.1}$$

where $\{\gamma^{(\mu)}\}$ is a sequence of positive numbers and L_1 is a convex-concave function defined by

$$L_1(z,p;\hat{p},\gamma) = \ell(z,p) - \frac{1}{2\gamma}|p-\hat{p}|^2, \tag{2.2}$$

where $|\cdot|$ denotes the Euclidean norm. Note that (2.1) means that the point $(z^{(\mu+1)}, p^{(\mu+1)})$ is an approximate solution of the min-max optimization problem on the right-hand side.

Since the function L_1 is quadratic in p for any fixed z, the inner maximization in (2.1) is equivalent to computing p by

$$p = p^{(\mu)} - \gamma^{(\mu)}(Az-b). \tag{2.3}$$

Substituting (2.3) into (2.1), we have

$$z^{(\mu+1)} \approx \arg\min_{z\in R^n} \phi_1(z;p^{(\mu)},\gamma^{(\mu)}), \tag{2.4}$$

where

$$\phi_1(z;\hat{p},\gamma) = \max_{p\in R^m} L_1(z,p;\hat{p},\gamma) = \ell(z,\hat{p}) + \frac{\gamma}{2}|Az-b|^2.$$

Note that the computation of p by (2.3) can be carried out exactly, while the minimization in (2.4) to compute $z^{(\mu+1)}$ can generally be carried out only approximately. The method of multipliers [1] may be stated as follows:

Algorithm MOM

Step 0: Let $\{\gamma^{(\mu)}\}$ be a sequence of positive numbers. Choose $p^{(0)}$ arbitrarily. Set $\mu := 0$.

Step 1: Compute $z^{(\mu+1)}$ by approximately minimizing $\phi_1(z; p^{(\mu)}, \gamma^{(\mu)})$.

Step 2: Let $p^{(\mu+1)} := p^{(\mu)} - \gamma^{(\mu)}(Az^{(\mu+1)} - b)$. Set $\mu := \mu + 1$ and go to Step 1.

Rockafellar [19, Theorem 4] shows the convergence of algorithm MOM under the following approximation criterion for (2.4):

$$\phi_1(z^{(\mu+1)}; p^{(\mu)}, \gamma^{(\mu)}) - \inf_z \phi_1(z; p^{(\mu)}, \gamma^{(\mu)}) \leq \frac{\left(\epsilon^{(\mu)}\right)^2}{2\gamma^{(\mu)}}, \tag{2.5}$$

where $\{\epsilon^{(\mu)}\}$ is a sequence of positive numbers such that $\sum_{\mu=0}^{\infty} \epsilon^{(\mu)} < \infty$. Generally speaking, it is difficult to check (2.5), because the exact minimum value $\inf_z \phi_1(z; p^{(\mu)}, \gamma^{(\mu)})$ is usually unknown.

On the other hand, assuming that the function F is strongly convex with modulus β, Kort and Bertsekas [16, Proposition 4] show the convergence of MOM under the following approximation criterion:

$$\mathrm{dist}(0, \partial\phi_1(z^{(\mu+1)}; p^{(\mu)}, \gamma^{(\mu)})) \leq \sqrt{\frac{\eta^{(\mu)}}{2\gamma^{(\mu)}}}, \tag{2.6}$$

where $\{\eta^{(\mu)}\}$ is a sequence of positive numbers such that $\eta^{(\mu)} < 2\beta$. In (2.6), $\partial\phi_1$ denotes the (set-valued) subdifferential of the convex function ϕ_1 and $\mathrm{dist}(0, S)$ denotes the distance between the origin and a set S. The criterion (2.6) is easier to check than (2.5) in that the former does not contain an unknown quantity.

The proximal method of multipliers (PMOM) is a variant of MOM, which generates a sequence $\{(z^{(\mu)}, p^{(\mu)})\}$ by the following iterative scheme:

$$(z^{(\mu+1)}, p^{(\mu+1)}) \approx \arg\min_{z \in R^n} \left\{ \max_{p \in R^m} L_2(z, p; z^{(\mu)}, p^{(\mu)}, \gamma^{(\mu)}) \right\}, \tag{2.7}$$

where $\{\gamma^{(\mu)}\}$ is a sequence of positive numbers and L_2 is a convex-concave function defined by

$$L_2(z, p; \hat{z}, \hat{p}, \gamma) = \ell(z, p) + \frac{1}{2\gamma}|z - \hat{z}|^2 - \frac{1}{2\gamma}|p - \hat{p}|^2. \tag{2.8}$$

Like the case of L_1, the exact maximizer of L_2 in p is given by (2.3). Thus, by (2.3) and (2.7), we have

$$z^{(\mu+1)} \approx \arg\min_{z \in R^n} \phi_2(z; z^{(\mu)}, p^{(\mu)}, \gamma^{(\mu)}),$$

where

$$\phi_2(z; \hat{z}, \hat{p}, \gamma) = \max_{p \in R^m} L_2(z, p; \hat{z}, \hat{p}, \gamma) = \ell(z, \hat{p}) + \frac{1}{2\gamma}|z - \hat{z}|^2 + \frac{\gamma}{2}|Az - b|^2.$$

To sum up, the proximal method of multipliers [19] may be stated as follows:

Algorithm PMOM

Step 0: Let $\{\gamma^{(\mu)}\}$ be a sequence of positive numbers. Choose $z^{(0)}$ and $p^{(0)}$ arbitrarily. Set $\mu := 0$.

Step 1: Compute $z^{(\mu+1)}$ by approximately minimizing $\phi_2(z; z^{(\mu)}, p^{(\mu)}, \gamma^{(\mu)})$.

Step 2: Let $p^{(\mu+1)} := p^{(\mu)} - \gamma^{(\mu)}(Az^{(\mu+1)} - b)$. Set $\mu := \mu + 1$ and go to Step 1.

The convergence of algorithm PMOM is established by Rockafellar [19, Theorem 7], under the following approximation criterion:

$$\mathrm{dist}(0, \partial\phi_2(z^{(\mu+1)}; z^{(\mu)}, p^{(\mu)}, \gamma^{(\mu)})) \leq \frac{\epsilon^{(\mu)}}{\gamma^{(\mu)}}, \tag{2.9}$$

where $\{\epsilon^{(\mu)}\}$ is a sequence of positive numbers such that $\sum_{\mu=0}^{\infty} \epsilon^{(\mu)} < \infty$. Like (2.6), criterion (2.9) does not contain an unknown quantity. Moreover, the strong convexity of the function F is not required unlike algorithm MOM with (2.6).

3. Algorithms

In this section, we shall focus our attention to problem (1.4). The Lagrangian ℓ for problem (1.4) may be written as

$$\ell(z, p) = \ell(x, y, p) = f(x) + g(y) - \langle p, A_1x + A_2y - b\rangle. \tag{3.1}$$

Let us consider the convex-concave function

$$L(x, y, p; \hat{x}, \hat{p}, \gamma) = \ell(x, y, p) + \frac{1}{2\gamma}|x - \hat{x}|^2 - \frac{1}{2\gamma}|p - \hat{p}|^2. \tag{3.2}$$

Notice the difference between this function and the functions L_1 and L_2 defined by (2.2) and (2.8), respectively. Since the function ℓ is strongly convex in y by assumption, the function L is strongly convex in (x, y).

Using the function L, we develop two algorithms that belong to the class of partial proximal method of multipliers (PPMOM). The first algorithm, called PPMOM$_1$, generates a sequence $\{(x^{(\mu)}, y^{(\mu)}, p^{(\mu)})\}$ by the following iterative scheme:

$$(x^{(\mu+1)}, y^{(\mu+1)}, p^{(\mu+1)}) \approx \arg \min_{(x,y)\in R^{n_1+n_2}} \left\{\max_{p\in R^m} L(x, y, p; x^{(\mu)}, p^{(\mu)}, \gamma^{(\mu)})\right\}, \tag{3.3}$$

where $\{\gamma^{(\mu)}\}$ is a sequence of positive numbers. Like (2.3), the exact maximizer of L is given by

$$p = p^{(\mu)} - \gamma^{(\mu)}(A_1x + A_2y - b),$$

so that

$$(x^{(\mu+1)}, y^{(\mu+1)}) \approx \arg \min_{(x,y)\in R^{n_1+n_2}} \phi(x, y; x^{(\mu)}, p^{(\mu)}, \gamma^{(\mu)}), \tag{3.4}$$

where

$$\begin{aligned}\phi(x, y; \hat{x}, \hat{p}, \gamma) &= \max_{p\in R^m} L(x, y, p; \hat{x}, \hat{p}, \gamma) \\ &= \ell(x, y, \hat{p}) + \frac{1}{2\gamma}|x - \hat{x}|^2 + \frac{\gamma}{2}|A_1x + A_2y - b|^2.\end{aligned}$$

Algorithm PPMOM$_1$ may be formally stated as follows:

Algorithm PPMOM$_1$

Step 0: Let $\{\gamma^{(\mu)}\}$ be a sequence of positive numbers. Choose $x^{(0)}$ and $p^{(0)}$ arbitrarily. Set $\mu := 0$.

Step 1: Compute $(x^{(\mu+1)}, y^{(\mu+1)})$ by approximately minimizing $\phi(x, y; x^{(\mu)}, p^{(\mu)}, \gamma^{(\mu)})$.

Step 2: Compute $p^{(\mu+1)} := p^{(\mu)} - \gamma^{(\mu)}(A_1x^{(\mu+1)} + A_2y^{(\mu+1)} - b)$. Set $\mu := \mu + 1$ and go to Step 1.

As an approximation criterion for the inexact minimization in (3.4), we use

$$\mathrm{dist}(0, \partial\phi(x^{(\mu+1)}, y^{(\mu+1)}; x^{(\mu)}, p^{(\mu)}, \gamma^{(\mu)})) \leq \frac{\beta^{(\mu)}\epsilon^{(\mu)}}{\gamma^{(\mu)}}, \tag{3.5}$$

where $\{\epsilon^{(\mu)}\}$ is a sequence of positive numbers such that $\sum_{\mu=0}^{\infty} \epsilon^{(\mu)} < \infty$ and $\beta^{(\mu)} = \min\{1, \beta\gamma^{(\mu)}\}$. ($\beta$ is the modulus of strong convexity of g.) The subdifferential of ϕ used in (3.5) is given by

$$\partial\phi(x, y; \hat{x}, \hat{p}, \gamma) = \partial_x\phi(x, y; \hat{x}, \hat{p}, \gamma) \times \partial_y\phi(x, y; \hat{x}, \hat{p}, \gamma) \tag{3.6}$$

with

$$\partial_x\phi(x, y; \hat{x}, \hat{p}, \gamma) = \partial_x\ell(x, y, \hat{p}) + \frac{1}{\gamma}(x - \hat{x}) + \gamma A_1^T(A_1x + A_2y - b) \tag{3.7}$$

and

$$\partial_y\phi(x, y; \hat{x}, \hat{p}, \gamma) = \partial_y\ell(x, y, \hat{p}) + \gamma A_2^T(A_1x + A_2y - b). \tag{3.8}$$

Note that, like (2.6) and (2.9), criterion (3.5) does not contain an unknown quantity such as the exact minimum value of a convex function. Note however that, owing to the quadratic term $(\gamma/2)|A_1x + A_2y - b|^2$, the function ϕ to be minimized in Step 1 does not enjoy the separability that the given problem possesses.

From this point of view, we propose another implementation of the partial proximal method of multipliers, which we call PPMOM$_2$:

$$(x^{(\mu+1)}, y^{(\mu+1)}, p^{(\mu+1)}) \approx \arg\max_{p\in R^m} \left\{ \min_{(x,y)\in R^{n_1+n_2}} L(x, y, p; x^{(\mu)}, p^{(\mu)}, \gamma^{(\mu)}) \right\}, \tag{3.9}$$

where $\{\gamma^{(\mu)}\}$ is a sequence of positive numbers. The difference between (3.9) and (3.3) consists in the order of min- and max-operations in computing the saddle point of L. The idea of reversing the order of min- and max-operations has also been considered by the authors [14] for the primal-dual proximal point algorithm.

Since the function L is separable in x and y, the inner minimization in (3.9) can be separately carried out in x and y. For any fixed p, let $(X(p;\ \hat{x}, \gamma), Y(p))$ be the exact minimizer of L, i.e.,

$$X(p;\ \hat{x}, \gamma) = \arg\min_{x\in R^{n_1}} \left\{ f(x) - \langle p, A_1x\rangle + \frac{1}{2\gamma}|x - \hat{x}|^2 \right\}, \tag{3.10}$$

$$Y(p) = \arg\min_{y\in R^{n_2}} \{g(y) - \langle p, A_2y\rangle\}. \tag{3.11}$$

Since L is strongly convex in (x, y) because of the partial proximal term $1/(2\gamma)|x - \hat{x}|^2$ and the strong convexity of g, $X(p;\ \hat{x}, \gamma)$ and $Y(p)$ are uniquely determined. Note that the minimizer $X(p;\ \hat{x},\gamma)$ in (3.10) depends on the iteration while $Y(p)$ in (3.11) does not. By (3.9), (3.10) and (3.11), we have

$$p^{(\mu+1)} \approx \arg\max_{p\in R^m} \psi(p; x^{(\mu)}, p^{(\mu)}, \gamma^{(\mu)}), \tag{3.12}$$

where

$$\begin{aligned}\psi(p;\ \hat{x},\hat{p},\gamma) &= \min_{(x,y)\in R^{n_1+n_2}} L(x,y,p;\ \hat{x},\hat{p},\gamma) \\ &= L(X(p;\ \hat{x},\gamma),\ Y(p),p;\ \hat{x},\hat{p},\gamma).\end{aligned} \tag{3.13}$$

The second version of the partial proximal method of multipliers, algorithm PPMOM$_2$, may then be stated as follows:

Algorithm PPMOM$_2$

Step 0: Let $\{\gamma^{(\mu)}\}$ be a sequence of positive numbers. Choose $x^{(0)}$ and $p^{(0)}$ arbitrarily. Set $\mu := 0$.

Step 1: Compute $p^{(\mu+1)}$ by approximately maximizing $\psi(p; x^{(\mu)}, p^{(\mu)}, \gamma^{(\mu)})$.

Step 2: Put $(x^{(\mu+1)}, y^{(\mu+1)}) := (X(p^{(\mu+1)};\ x^{(\mu)}, \gamma^{(\mu)}), Y(p^{(\mu+1)}))$. Set $\mu := \mu+1$ and go to Step 1.

Note, in particular, that the function ψ defined by (3.13) is differentiable and its gradient is given by

$$\nabla\psi(p;\ \hat{x},\hat{p},\gamma) = -\frac{1}{\gamma}(p-\hat{p}) - (A_1X(p;\ \hat{x},\gamma) + A_2Y(p) - b). \tag{3.14}$$

We use the following approximation criterion for the inexact maximization in (3.12):

$$|\nabla\psi(p^{(\mu+1)}; x^{(\mu)}, p^{(\mu)}, \gamma^{(\mu)})| \le \frac{\beta^{(\mu)}\epsilon^{(\mu)}}{\gamma^{(\mu)}}, \tag{3.15}$$

where $\{\epsilon^{(\mu)}\}$ and $\{\beta^{(\mu)}\}$ are the same as in (3.5). The convergence of these two algorithms will be established in Section 5.

4. Basic results

A pair of optimal solutions (x^*, y^*) and p^* to problem (1.4) and its dual, respectively, satisfies the Kuhn-Tucker conditions

$$0 \in \partial f(x^*) - A_1^T p^*, \quad 0 \in \partial g(y^*) - A_2^T p^*, \quad A_1x^* + A_2y^* = b. \tag{4.1}$$

In addition, (x^*, y^*, p^*) is a saddle point of the Lagrangian function ℓ defined by (3.1). Suppose that problem (1.4) has an optimal solution (x^*, y^*) and satisfies the constraint qualification

$$\mathrm{ri}(\mathrm{dom}(f) \times \mathrm{dom}(g)) \cap \{(x,y)|A_1x + A_2y = b\} \neq \emptyset,$$

where ri$(\cdot)$ and dom$(\cdot)$ denote the relative interior of a convex set and the effective domain of a convex function, respectively. Then there exists a vector p^* satisfying (4.1) [18, Corollary 29.1.4].

Associated with the Lagrangian ℓ is a point-to-set mapping $T_\ell : R^{n_1+n_2+m} \to R^{n_1+n_2+m}$ defined by

$$\begin{aligned} T_\ell(x,y,p) &= \{(u,v,s) \mid u \in \partial_x \ell(x,y,p),\ \ v \in \partial_y \ell(x,y,p),\ \ s \in -\partial_p \ell(x,y,p)\} \\ &= \{(u,v,s) \mid u \in \partial f(x) - A_1^T p,\ \ v \in \partial g(y) - A_2^T p,\ \ s = A_1 x + A_2 y - b\} \end{aligned} \quad (4.2)$$

Since the convex-concave function ℓ is closed and proper [18, pp. 362-363], the mapping T_ℓ is maximal monotone [18, Corollary 37.5.2]. In view of the definition (4.2) of T_ℓ, the zeros of T_ℓ satisfy the Kuhn-Tucker conditions (4.1), and hence they solve problem (1.4) and its dual. Notice that since the function g is strongly convex with modulus β, T_ℓ is strongly monotone with modulus β with respect to the second component y, i.e.,

$$\langle (x,y,p) - (x',y',p'),\ \ (u,v,s) - (u',v',s') \rangle \geq \beta |y - y'|^2, \quad (4.3)$$

for all $(u,v,s) \in T_\ell(x,y,p)$ and $(u',v',s') \in T_\ell(x',y',p')$.

Consider the following parametrized problem perturbed by $(u,v,s) \in R^{n_1+n_2+m}$:

$$\begin{array}{ll} \underset{(x,y)\in R^{n_1+n_2}}{\text{minimize}} & f(x) + g(y) - \langle u, x \rangle - \langle v, y \rangle \\ \text{subject to} & A_1 x + A_2 y = b + s. \end{array} \quad (4.4)$$

Then we can show the next result.

Proposition 1 *Suppose that, for some constant $\lambda > 0$, the parametrized problem (4.4) has an optimal solution, whenever $\max\{|u|, |v|, |s|\} < \lambda$. Then we have*

$$0 \in \text{int im}(T_\ell), \quad (4.5)$$

where $\text{int}(\cdot)$ and $\text{im}(\cdot)$ denote the interior of a convex set and the image of a mapping, respectively.

Proof. Let $V(s)$ denote the set $\{(x,y) \mid A_1 x + A_2 y = b + s\}$. It suffices to show that

$$\text{ri}(\text{dom}(f) \times \text{dom}(g)) \cap V(s) \neq \emptyset, \quad (4.6)$$

for any vector s such that $|s| < \lambda$. Because, then, for each (u,v,s) such that $\max\{|u|, |v|, |s|\} < \lambda$, there exists a Lagrange multiplier vector p associated with an optimal solution (x,y) of (4.4) satisfying

$$u \in \partial_x \ell(x,y,p),\ \ v \in \partial_y \ell(x,y,p),\ \ s \in -\partial_p \ell(x,y,p),$$

i.e.,

$$(u,v,s) \in T_\ell(x,y,p),$$

which implies (4.5).

To show that (4.6) holds for any s such that $|s| < \lambda$, we assume to the contrary that there exists some vector s' such that $|s'| < \lambda$ and

$$\text{ri}(\text{dom}(f) \times \text{dom}(g)) \cap V(s') = \emptyset.$$

Under the hypothesis of the proposition, it is obvious that

$$(\text{dom}(f) \times \text{dom}(g)) \cap V(s) \neq \emptyset \quad (4.7)$$

for any vector s satisfying $|s| < \lambda$. Therefore, we have

$$(\mathrm{dom}(f) \times \mathrm{dom}(g)) \cap V(s') \subset \mathrm{rb}(\mathrm{dom}(f) \times \mathrm{dom}(g)),$$

where $\mathrm{rb}(\cdot)$ denotes the relative boundary of the set. Then there exists a hyperplane $H \subset R^{n_1+n_2}$ containing $V(s')$ such that $(\mathrm{dom}(f) \times \mathrm{dom}(g)) \subset H^+$, where H^+ is a half space defined by H. Let H^- be another half space defined by H and choose $(\hat{x}, \hat{y}) \in \mathrm{int}(H^-)$ arbitrarily. Then, letting $\hat{s} = A_1\hat{x} + A_2\hat{y} - b \in R^m$, we have $V(s' + \delta(\hat{s} - s')) \subset \mathrm{int}(H^-)$ for any $\delta > 0$. Since $(\mathrm{dom}(f) \times \mathrm{dom}(g)) \subset H^+$, this implies that, for all $\delta > 0$ small enough,

$$(\mathrm{dom}(f) \times \mathrm{dom}(g)) \cap V(s' + \delta(\hat{s} - s')) = \emptyset.$$

Since $|s'| < \lambda$, this contradicts (4.7) and the proof is complete. □

The hypothesis of Proposition 1 can be regarded as a constraint qualification for problem (1.4). We may expect that it usually, if not always, holds when the original problem (1.4) has a solution and the feasible set of problem (1.4) has a nonempty intersection with $\mathrm{dom}(f) \times \mathrm{dom}(g)$ provided that the perturbation s is small enough.

Recall that each iteration of both algorithms PPMOM_1 and PPMOM_2 consists of finding a saddle point of the function L defined by (3.2). Any saddle point (x, y, p) of L satisfies

$$0 \in \partial L(x, y, p;\ \hat{x}, \hat{p}, \gamma) = \partial \ell(x, y, p) + \frac{1}{\gamma}(x - \hat{x}, 0, -(p - \hat{p})).$$

From (4.2), it follows that

$$(\hat{x}, 0, \hat{p}) \in (x, 0, p) + \gamma T_\ell(x, y, p). \tag{4.8}$$

Let $\Pi : R^{n_1+n_2+m} \to R^{n_1+n_2+m}$ denote the projection mapping onto the space of variables x and p, i.e.,

$$\Pi(x, y, p) = (x, 0, p), \quad \forall (x, y, p) \in R^{n_1+n_2+m}.$$

Then (4.8) can be written as

$$\Pi(\hat{x}, \hat{y}, \hat{p}) \in (\Pi + \gamma T_\ell)(x, y, p),$$

or equivalently,

$$(x, y, p) \in (\Pi + \gamma T_\ell)^{-1}\Pi(\hat{x}, \hat{y}, \hat{p}). \tag{4.9}$$

Thus the set of saddle points of the function L may be formally expressed as the right-hand side of (4.9).

We show some properties of the mapping $Q = (\Pi + \gamma T_\ell)^{-1}$ in Proposition 2, and then show some properties of $P = Q\Pi = (\Pi + \gamma T_\ell)^{-1}\Pi$ in Proposition 3.

Proposition 2 *Let* $Q = (\Pi + \gamma T_\ell)^{-1}$. *Suppose that* $\mathrm{dom}(T_\ell) \neq \emptyset$. *Then we have the following:*
(i) The mapping Q *is single-valued on* $R^{n_1+n_2+m}$.
(ii) For any $(x, y, p), (x', y', p') \in R^{n_1+n_2+m}$,

$$|Q(x, y, p) - Q(x', y', p')| \leq \frac{1}{\beta'}|(x, y, p) - (x', y', p')|, \tag{4.10}$$

where $\beta' = \min\{1, \beta\gamma\}$.

Proof. (i) By [5, Theorem 2.7], it is sufficient to show that the mapping $(\Pi+\gamma T_\ell)$ is maximal monotone and coercive.

First, we derive the maximal monotonicity of $(\Pi + \gamma T_\ell)$. By [5, Proposition 2.10], γT_ℓ is maximal monotone for any $\gamma > 0$, since so is T_ℓ. Therefore, by [5, Theorem 2.3], the mapping $(\Pi + \gamma T_\ell)$ is maximal monotone, because

$$\mathrm{dom}(\gamma T_\ell) \cap \mathrm{int}\ \mathrm{dom}(\Pi) = \mathrm{dom}(T_\ell) \cap R^{n_1+n_2+m} \neq \emptyset.$$

Next we show that $(\Pi + \gamma T_\ell)$ is strongly monotone. For any $(u, v, s) \in T_\ell(x, y, p)$ and $(u', v', s') \in T_\ell(x', y', p')$, we have

$$(x, 0, p) + \gamma(u, v, s) \in (\Pi + \gamma T_\ell)(x, y, p)$$

and

$$(x', 0, p') + \gamma(u', v', s') \in (\Pi + \gamma T_\ell)(x', y', p'),$$

respectively. Then, by (4.3), we have

$$\begin{aligned}
&\langle (x, y, p) - (x', y', p'),\ \ (x, 0, p) + \gamma(u, v, s) - \{(x', 0, p') + \gamma(u', v', s')\}\rangle \\
&\quad = |x - x'|^2 + |p - p'|^2 + \gamma\langle (x, y, p) - (x', y', p'),\ \ (u, v, s) - (u', v', s')\rangle \\
&\quad \geq |x - x'|^2 + |p - p'|^2 + \beta\gamma|y - y'|^2 \\
&\quad \geq \beta'|(x, y, p) - (x', y', p')|^2,
\end{aligned}$$

where $\beta' = \min\{1, \beta\gamma\}$. This implies that $(\Pi + \gamma T_\ell)$ is strongly monotone. Since the strong monotonicity implies the coerciveness, we have proved (i).

(ii) For any $(x, y, p) \in R^{n_1+n_2+m}$, let $(x_+, y_+, p_+) = Q(x, y, p)$. (Note that the existence and the single-valuedness of $Q(x, y, p)$ are assured by (i).) Then, it follows from the definition of Q that

$$(x, y, p) \in (\Pi + \gamma T_\ell)(x_+, y_+, p_+),$$

or

$$(x, y, p) = (x_+, 0, p_+) + \gamma(u_+, v_+, s_+) \tag{4.11}$$

for some vector $(u_+, v_+, s_+) \in T_\ell(x_+, y_+, p_+)$. Similarly, for any $(x', y', p') \in R^{n_1+n_2+m}$, we have

$$(x', y', p') = (x'_+, 0, p'_+) + \gamma(u'_+, v'_+, s'_+), \tag{4.12}$$

where $(x'_+, y'_+, p'_+) = Q(x', y', p')$ and $(u'_+, v'_+, s'_+) \in T_\ell(x'_+, y'_+, p'_+)$. By (4.11) and (4.12), we have

$$(u_+, v_+, s_+) - (u'_+, v'_+, s'_+) = \frac{1}{\gamma}\{(x, y, p) - (x', y', p') - (x_+, 0, p_+) + (x'_+, 0, p'_+)\}. \tag{4.13}$$

Since T_ℓ is strongly monotone with respect to the second component (c.f. (4.3)), we have

$$\langle (u_+, v_+, s_+) - (u'_+, v'_+, s'_+),\ \ (x_+, y_+, p_+) - (x'_+, y'_+, p'_+)\rangle \geq \beta|y_+ - y'_+|^2. \tag{4.14}$$

It follows from (4.13) and (4.14) that

$$\begin{aligned}
&\langle (x, y, p) - (x', y', p'),\ \ (x_+, y_+, p_+) - (x'_+, y'_+, p'_+)\rangle \\
&\qquad \geq\ |x_+ - x'_+|^2 + \beta\gamma|y_+ - y'_+|^2 + |p_+ - p'_+|^2 \\
&\qquad \geq\ \beta'|(x_+, y_+, p_+) - (x'_+, y'_+, p'_+)|^2,
\end{aligned}$$

where $\beta' = \min\{1, \beta\gamma\}$. Therefore we have

$$\begin{aligned} |(x,y,p) - (x',y',p')| &\geq \beta'|(x_+,y_+,p_+) - (x'_+,y'_+,p'_+)| \\ &= \beta'|Q(x,y,p) - Q(x',y',p')|, \end{aligned}$$

which completes the proof. □

From Proposition 2, we obtain the following results immediately.

Proposition 3 *Let* $P = Q\Pi = (\Pi + \gamma T_\ell)^{-1}\Pi$. *Suppose that* $\mathrm{dom}(T_\ell) \neq \emptyset$. *Then we have the following:*
(i) The mapping P *is single-valued on* $R^{n_1+n_2+m}$.
(ii) For any $(x,y,p), (x',y',p') \in R^{n_1+n_2+m}$,

$$|P(x,y,p) - P(x',y',p')| \leq \frac{1}{\beta'}|(x,p) - (x',p')|, \tag{4.15}$$

where $\beta' = \min\{1, \beta\gamma\}$.

Proof. (i) Obvious from Proposition 2 (i).
(ii) From (4.10), we have

$$\begin{aligned} |P(x,y,p) - P(x',y',p')| &= |Q\Pi(x,y,p) - Q\Pi(x',y',p')| \\ &\leq \frac{1}{\beta'}|\Pi(x,y,p) - \Pi(x',y',p')| \\ &= \frac{1}{\beta'}|(x,p) - (x',p')|. \end{aligned}$$

This completes the proof. □

As mentioned above, each iteration of algorithms PPMOM_1 and PPMOM_2 consists of computing a saddle point of the function L, and hence the formulas (3.3) and (3.9) are regarded as particular realizations of the scheme

$$(x^{(\mu+1)}, y^{(\mu+1)}, p^{(\mu+1)}) \approx P^{(\mu)}(x^{(\mu)}, y^{(\mu)}, p^{(\mu)}), \tag{4.16}$$

where $P^{(\mu)}$ is defined by

$$P^{(\mu)} = (\Pi + \gamma^{(\mu)}T_\ell)^{-1}\Pi. \tag{4.17}$$

Note that the mapping $P^{(\mu)}$ is single-valued by Proposition 3. If Π is replaced by the identity mapping I in (4.17), the iteration (4.16) becomes

$$(x^{(\mu+1)}, y^{(\mu+1)}, p^{(\mu+1)}) \approx P^{(\mu)}(x^{(\mu)}, y^{(\mu)}, p^{(\mu)}) = (I + \gamma^{(\mu)}T_\ell)^{-1}(x^{(\mu)}, y^{(\mu)}, p^{(\mu)}).$$

This is nothing but the proximal method of multipliers, of which particular realization is algorithm PMOM described in Section 2.

5. Convergence

5.1. Convergence of algorithm PPMOM_1

The purpose of this section is to establish a convergence theorem for algorithm PPMOM_1. First, we show the following proposition.

Proposition 4 *For any* $(x,y,p) \in R^{n_1+n_2+m}$*, we have*

$$|(x,y,p) - P(\hat{x},\hat{y},\hat{p})| \leq \frac{\gamma}{\beta'}\mathrm{dist}(0, \partial\phi(x,y;\ \hat{x},\hat{p},\gamma)), \tag{5.1}$$

where $P = Q\Pi = (\Pi + \gamma T_\ell)^{-1}\Pi$ *and* $\partial\phi$ *is given by* (3.6)-(3.8).

Proof. Let us choose an arbitrary vector $\omega = (\omega_1, \omega_2) \in \partial\phi(x,y;\ \hat{x},\hat{p},\gamma)$. From (3.6)-(3.8), we have

$$\begin{aligned} \omega_1 + \frac{1}{\gamma}(\hat{x} - x) &\in \partial_x \ell(x,y,\hat{p}) + \gamma A_1^T(A_1x + A_2y - b) \\ &= \partial f(x) - A_1^T p \end{aligned} \tag{5.2}$$

and

$$\begin{aligned} \omega_2 &\in \partial_y \ell(x,y,\hat{p}) + \gamma A_2^T(A_1x + A_2y - b) \\ &= \partial g(y) - A_2^T p, \end{aligned} \tag{5.3}$$

where

$$p = \hat{p} - \gamma(A_1x + A_2y - b). \tag{5.4}$$

In view of the definition (4.2) of T_ℓ, (5.2)-(5.4) are written as

$$(\gamma\omega_1 + \hat{x} - x,\ \gamma\omega_2,\ \hat{p} - p) \in \gamma T_\ell(x,y,p).$$

Therefore, we have

$$(\gamma\omega_1 + \hat{x},\ \gamma\omega_2,\ \hat{p}) \in (\Pi + \gamma T_\ell)(x,y,p),$$

or equivalently,

$$(x,y,p) = Q(\gamma\omega_1 + \hat{x},\ \gamma\omega_2,\ \hat{p}), \tag{5.5}$$

where $Q = (\Pi + \gamma T_\ell)^{-1}$. (Note that Q is single-valued by Proposition 2.) From (5.5), it follows that

$$|(x,y,p) - P(\hat{x},\hat{y},\hat{p})| = |(x,y,p) - Q\Pi(\hat{x},\hat{y},\hat{p})| = |Q(\gamma\omega_1 + \hat{x},\ \gamma\omega_2,\ \hat{p}) - Q(\hat{x},0,\hat{p})|.$$

By Proposition 2, we have

$$\begin{aligned} |Q(\gamma\omega_1 + \hat{x},\ \gamma\omega_2,\ \hat{p}) - Q(\hat{x},0,\hat{p})| &\leq \frac{1}{\beta'}|(\gamma\omega_1 + \hat{x},\ \gamma\omega_2,\ \hat{p}) - (\hat{x},0,\hat{p})| \\ &= \frac{\gamma}{\beta'}|\omega|. \end{aligned}$$

Since $\omega \in \partial\phi(x,y;\ \hat{x},\hat{p},\gamma)$ was arbitrary, we obtain (5.1). □

Now, we are ready to show the following convergence theorem for algorithm PPMOM_1.

Theorem 1 *Suppose that the hypotheses of Proposition 1 are satisfied. Then the sequence* $\{(x^{(\mu)}, y^{(\mu)}, p^{(\mu)})\}$ *generated by algorithm* PPMOM_1 *with* (3.5) *converges to a vector* (x^*, y^*, p^*) *satisfying the Kuhn-Tucker conditions* (4.1) *for problem* (1.4).

Proof. Under the given hypotheses, there exists an optimal solution (x^*, y^*) of problem (1.4) together with a Lagrange multiplier vector p^*, satisfying the Kuhn-Tucker conditions (4.1). Hence it follows from Proposition 1 that $0 \in \text{int im}(T_\ell)$. Furthermore, by Proposition 4, the approximation criterion (3.5) implies

$$|(x^{(\mu+1)}, y^{(\mu+1)}, p^{(\mu+1)}) - P^{(\mu)}(x^{(\mu)}, y^{(\mu)}, p^{(\mu)})| \leq \epsilon^{(\mu)}.$$

Then, from Theorem 1 of [12], any of the limit points of the sequence $\{(x^{(\mu)}, y^{(\mu)}, p^{(\mu)})\}$ is a zero of T_ℓ and $\{(x^{(\mu)}, p^{(\mu)})\}$ converges. Since g is strongly convex, the y-component of a solution of (1.4) is uniquely determined, which implies that $\{y^{(\mu)}\}$ converges. Thus the whole sequence $\{(x^{(\mu)}, y^{(\mu)}, p^{(\mu)})\}$ converges to a solution of (1.4). □

5.2. Convergence of algorithm PPMOM$_2$

In this section, we establish a convergence theorem for algorithm PPMOM$_2$. We first show the following proposition.

Proposition 5 *For any $(x, y, p) \in R^{n_1+n_2+m}$, we have*

$$|(x, y, p) - P(\hat{x}, \hat{y}, \hat{p})| \leq \frac{\gamma}{\beta'}|\nabla\psi(p;\ \hat{x}, \hat{p}, \gamma)|, \tag{5.6}$$

where $P = (\Pi + \gamma T_\ell)^{-1}\Pi$ and $\nabla\psi$ is given by (3.14).

Proof. Put, for convenience, $\xi = \nabla\psi(p;\ \hat{x}, \hat{p}, \gamma)$. From (3.14), we have

$$\xi = -\frac{1}{\gamma}(p - \hat{p}) - (A_1 x + A_2 y - b),$$

where $x = X(p;\ \hat{x}, \gamma)$ and $y = Y(p)$. Then, it follows from (3.1) that

$$\xi + \frac{1}{\gamma}(p - \hat{p}) = -(A_1 x + A_2 y - b) \in \partial_p \ell(x, y, p). \tag{5.7}$$

In view of (3.10) and (3.11), it holds that

$$\frac{1}{\gamma}(\hat{x} - x) \in \partial_x \ell(x, y, p), \quad 0 \in \partial_y \ell(x, y, p). \tag{5.8}$$

From (5.7) and (5.8), we have

$$(\hat{x} - x,\ 0,\ \hat{p} - p - \gamma\xi) \in \gamma T_\ell(x, y, p),$$

which implies that

$$(x, y, p) = P(\hat{x},\ \hat{y},\ \hat{p} - \gamma\xi).$$

Therefore, it holds that

$$|(x, y, p) - P(\hat{x}, \hat{y}, \hat{p})| = |P(\hat{x}, \hat{y}, \hat{p} - \gamma\xi) - P(\hat{x}, \hat{y}, \hat{p})|. \tag{5.9}$$

By Proposition 3, we have

$$|P(\hat{x}, \hat{y}, \hat{p} - \gamma\xi) - P(\hat{x}, \hat{y}, \hat{p})| \leq \frac{1}{\beta'}|(\hat{x}, \hat{p} - \gamma\xi) - (\hat{x}, \hat{p})| = \frac{\gamma}{\beta'}|\xi|. \tag{5.10}$$

From (5.9) and (5.10), we obtain (5.6). □

Now we state a convergence theorem for algorithm PPMOM$_2$.

Theorem 2 *Suppose that the hypotheses of Proposition 1 are satisfied. Then the sequence* $\{(x^{(\mu)}, y^{(\mu)}, p^{(\mu)})\}$ *generated by algorithm* $PPMOM_2$ *with* (3.15) *converges to a vector* (x^*, y^*, p^*) *satisfying the Kuhn-Tucker conditions* (4.1) *for problem* (1.4).

Proof. The proof is similar to Theorem 1. By Proposition 1, we obtain $0 \in \text{int im}(T_\ell)$ and, by Proposition 5,

$$|(x^{(\mu+1)}, y^{(\mu+1)}, p^{(\mu+1)}) - P^{(\mu)}(x^{(\mu)}, y^{(\mu)}, p^{(\mu)})| \leq \epsilon^{(\mu)}.$$

By using Theorem 1 of [12], we obtain the desired result. □

6. Application to problems with multiple set constraints

In this section, we consider the following convex programming problem:

$$\begin{array}{ll} \text{minimize} & g(y) \\ \text{subject to} & y \in C_1 \cap C_2 \cap \cdots \cap C_N, \end{array} \tag{6.1}$$

where the function $g : R^n \to R \cup \{\infty\}$ is closed, proper and strongly convex (with modulus β) and the sets $C_1, C_2, \ldots, C_N \subset R^n$ are closed and convex. We suppose that each C_i is individually so simple that the projection onto C_i is easy to compute. We shall show how algorithm $PPMOM_2$ can effectively be applied to problem (6.1).

Let us reformulate problem (6.1) in the form (1.4) as follows: For each i, let $f_i : R^n \to R \cup \{+\infty\}$ be the indicator function of C_i, namely

$$f_i(x_i) = \begin{cases} 0, & \text{if } x_i \in C_i, \\ +\infty, & \text{otherwise.} \end{cases}$$

Then problem (6.1) can be written as

$$\begin{array}{ll} \text{minimize} & \sum_{i=1}^{N} f_i(x_i) + g(y) \\ \text{subject to} & x_1 = x_2 = \cdots = x_N = y. \end{array} \tag{6.2}$$

Now let A_1 be the $n \times n$ identity matrix I, A_2 be the $Nn \times n$ matrix $[-I, -I, \cdots, -I]^T$ and $b = 0$. Furthermore, let $f : R^{Nn} \to R \cup \{+\infty\}$ be given by

$$f(x) = \sum_{i=1}^{N} f_i(x_i).$$

Then, we see that problem (6.2) is represented as problem (1.4). Note in particular that the function f has a separable structure.

It follows from (3.10) that

$$X(p;\ \hat{x}, \gamma) = \arg \min_{x \in R^{Nn}} \left\{ f(x) - \langle p, x \rangle + \frac{1}{2\gamma}|x - \hat{x}|^2 \right\}.$$

Thanks to the separability of f, each component of $X(p;\ \hat{x}, \gamma)$ can be computed individually as

$$\begin{aligned} X_i(p_i;\ \hat{x}_i, \gamma) &= \arg \min_{x_i \in R^n} \left\{ f_i(x_i) - \langle p_i, x_i \rangle + \frac{1}{2\gamma}|x_i - \hat{x}_i|^2 \right\} \\ &= \text{Proj}_{C_i}(\hat{x}_i + \gamma p_i), \end{aligned}$$

where p_i and $\hat{x}_i$ are the ith components of p and $\hat{x}$, respectively, and $\mathrm{Proj}_{C_i}(\cdot)$ denotes the projection onto the set C_i.

On the other hand, we have from (3.11) that

$$Y(p) = \arg\min_{y \in R^n} \left\{ g(y) + \sum_{i=1}^{N} \langle p_i, y \rangle \right\}.$$

Thus $Y(p)$ is uniquely obtained by minimizing the strongly convex function subject to no constraints. If the function g is separable in y_j, $j = 1, 2, \ldots, n$, then $Y(p)$ can be computed componentwise. With $X(p;\ \hat{x}, \gamma)$ and $Y(p)$ thus computed, the value of $\psi(p;\ \hat{x}, \hat{p}, \gamma)$ and its gradient $\nabla\psi(p;\ \hat{x}, \hat{p}, \gamma)$ can be evaluated by (3.13) and (3.14), respectively. Therefore, the maximization of ψ in Step 1 of algorithm PPMOM$_2$ may be carried out using any gradient-type algorithm such as quasi-Newton methods.

7. Application to traffic assignment problems

In this section, we apply algorithm PPMOM$_2$ to traffic assignment problems. Consider a directed network $G = (\mathcal{N}, \mathcal{A})$, where $\mathcal{N}$ and $\mathcal{A}$ are the sets of nodes and arcs, respectively. Let

K: number of commodities (O-D pairs),
$x_{kj} \in R$: flow of commodity k on arc j,
$y_j \in R$: total flow on arc j, i.e., $y_j = \sum_{k=1}^{K} x_{kj}$,
$g_j : R \to R \cup \{+\infty\}$: travel cost on arc j dependent on total flow y_j,
$O(i) \subset \mathcal{A}$: set of arcs originating at node i,
$I(i) \subset \mathcal{A}$: set of arcs terminating at node i,
$d_{ki} \in R$: demand for commodity k at node i.

Then the traffic assignment problem is formulated as follows:

$$\begin{array}{ll} \text{minimize} & \displaystyle\sum_{j \in \mathcal{A}} g_j(y_j) \\ \text{subject to} & y_j = \displaystyle\sum_{k=1}^{K} x_{kj}, \quad j \in \mathcal{A} \\ & \displaystyle\sum_{j \in O(i)} x_{kj} - \sum_{j \in I(i)} x_{kj} = d_{ki}, \quad k = 1, 2, \ldots, K, \quad i \in \mathcal{N} \\ & x_{kj} \geq 0, \quad k = 1, 2, \ldots, K, \quad j \in \mathcal{A}. \end{array} \tag{7.1}$$

Note that the solution of problem (7.1) corresponds to the Wardrop's user optimal equilibrium [17].

Let $f_k : R^{|\mathcal{A}|} \to R \cup \{+\infty\}$ be defined by

$$f_k(x_k) = \begin{cases} 0, & \text{if } Ex_k = d_k \text{ and } x_k \geq 0, \\ +\infty, & \text{otherwise,} \end{cases}$$

where $E \in R^{|\mathcal{N}| \times |\mathcal{A}|}$ denotes the node-arc incidence matrix of $G = (\mathcal{N}, \mathcal{A})$, $x_k = (x_{k1}, \ldots, x_{k|\mathcal{A}|})^T$ and $d_k = (d_{k1}, \ldots, d_{k|\mathcal{N}|})^T$. Then problem (7.1) is expressed as follows:

$$\begin{array}{ll} \text{minimize} & \displaystyle\sum_{k=1}^{K} f_k(x_k) + \sum_{j \in \mathcal{A}} g_j(y_j) \\ \text{subject to} & y_j = \displaystyle\sum_{k=1}^{K} x_{kj}, \quad j \in \mathcal{A}. \end{array} \tag{7.2}$$

Furthermore, problem (7.2) can be reformulated as problem (1.4) in Section 1 with the following identifications:

$$A_1 = [I,\ I, \cdots, I]\,,\ \ A_2 = -I,\ b = 0,$$

$$f(x) = \sum_{k=1}^{K} f_k(x_k)$$

and

$$g(y) = \sum_{j\in\mathcal{A}} g_j(y_j),$$

where I is the $|\mathcal{A}| \times |\mathcal{A}|$ identity matrix.

Assuming that the functions g_j, $j \in \mathcal{A}$, are strongly convex with modulus β, we consider applying algorithm PPMOM$_2$ to the traffic assignment problem (7.1). Since f is separable with respect to commodities k, $X(p;\ \hat{x}, \gamma)$ defined by (3.10) can be computed componentwise as

$$\begin{aligned} X_k(p;\ \hat{x}_k, \gamma) &= \arg \min_{x_k \in R^{|\mathcal{A}|}} \left\{ f_k(x_k) - \langle p, x_k\rangle + \frac{1}{2\gamma}|x_k - \hat{x}_k|^2 \right\} \\ &= \arg \min_{Ex_k = d_k, x_k \ge 0} \left\{ \frac{1}{2\gamma}|x_k - \hat{x}_k|^2 - \langle p, x_k\rangle \right\},\quad k = 1, \dots, K, \end{aligned}$$

which give rise to K independent single-commodity network flow problems with quadratic cost. Similarly, $Y(p)$ defined by (3.11) can also be evaluated componentwise by solving $|\mathcal{A}|$ independent univariate minimization problems, i.e.,

$$Y_j(p_j) = \arg \min_{y_j \in R} \{g_j(y_j) + p_j y_j\},\quad j \in \mathcal{A}.$$

8. Concluding remarks

For problem (1.4), we have proposed two variants of the partial proximal method of multipliers (PPMOM$_1$ and PPMOM$_2$), which compute in each iteration an approximate saddle point of the argumented Lagrangian L. The difference between PPMOM$_1$ and PPMOM$_2$ lies in the order of the (x, y)-minimization and the p-maximization in finding an approximate saddle point. In PPMOM$_1$, the p-maximization of L with (x, y) being fixed is easy to compute, but the resulting (x, y)-minimization problem becomes somewhat complicated. In PPMOM$_2$, the (x, y)-minimization is separately carried out in x and y and the resulting p-maximization problem becomes a differentiable optimization problem. Moreover, when the functions f and g are in particular separable, both the x- and y-minimizations are done in parallel (see Section 7).

As briefly mentioned in Introduction, some researchers have studied the partial proximal method. In [3], Bertsekas and Tseng extensively studied the primal version of the partial proximal method for convex programming problems. They also consider the dual version of the partial proximal method and discuss its relation with a decomposition method in convex programming. On the other hand, Ha [12] established some convergence results for the partial proximal method. The latter paper, however, only considers a basic iterative method under the general framework of maximal monotone mappings. It does not give any concrete procedure like PPMOM$_1$ and PPMOM$_2$, which take into account a particular structure of the problem to be solved.

Finally, we mention some other methods related to the algorithms proposed in this paper. For the special case of (1.4), where $A_2 = -I$ and $b = 0$, the alternating direction method of multipliers (ADMOM) have been proposed by Gabay [11] (see also [2, 6, 7]). Each iteration of algorithm ADMOM consists of three operations; the x-minimization and the y-minimization of the augmented Lagrangian, followed by the update of the multipiers p. ADMOM has turned out to be a useful decomposition algorithm for problems with some separable structure [8, 9]. Tseng [20] proposed an alternating minimization algorithm similar to ADMOM for solving (1.4), in which the function g is also supposed to be strongly convex. This method differs from ADMOM in that the y-minimization is done with respect to the ordinary Lagrangian and hence, like PPMOM_2, makes the most of the separability of function g the original problem may have.

Acknowledgement This work was supported in part by the Scientific Research Grant-in-Aid from the Ministry of Education, Science and Culture, Japan (No. 06750417 and No. 06650443).

References

[1] Bertsekas, D.P.: *Constrained Optimization and Lagrange Multiplier Methods.* Academic Press, New York, 1982.

[2] Bertsekas, D.P. and Tsitsiklis, J.N.: *Parallel and Distributed Computation: Numerical Methods.* Prentice Hall, New Jersey, 1989.

[3] Bertsekas, D.P. and Tseng, P.: "Partial proximal minimization algorithms for convex programming," *SIAM Journal on Optimization*, Vol. 4 (1994), 551-572.

[4] Chen, G. and Teboulle, M.: "A proximal-based decomposition method for convex minimization problems," *Mathematical Programming*, Vol. 64 (1992), 81-101.

[5] Dolezal, V.: *Monotone Operators and Applications in Control and Network Theory.* Elsevier, Amsterdam, 1979.

[6] Eckstein, J.: "Parallel alternating direction multiplier decomposition of convex programs," *Journal of Optimization Theory and Applications*, Vol. 80 (1994), 39-62.

[7] Eckstein, J. and Bertsekas, D.P.: "On the Douglas-Rachford splitting method and the proximal point algorithm for maximal monotone operators," *Mathematical Programming*, Vol. 55 (1992), 293-318.

[8] Eckstein, J. and Fukushima, M.: "Some reformulations and applications of the alternating direction method of multipliers," in W.W. Hager, D.W. Hearn and P.M. Pardalos, eds., *Large Scale Optimization: State of the Art* (Kluwer, Boston, 1994), 115-134.

[9] Fukushima, M.: "The primal Douglas-Rachford splitting algorithm for a class of monotone mappings with application to the traffic equilibrium problem," *Mathematical Programming*, to appear.

[10] Fukushima, M., Haddou, M., Nguyen, V.H., Strodiot, J.-J., Sugimoto, T. and Yamakawa, E.: "A parallel descent algorithm for convex programming," *Computational Optimization and Applications*, to appear.

[11] Gabay, D.: "Applications of the method of multipliers to variational inequalities," in: M. Fortin and R. Glowinski, eds., *Augmented Lagrangian Methods: Applications to the Numerical Solution of Boundary-Value Problems* (North-Holland, Amsterdam, 1983), 299-331.

[12] Ha, C.D.: "A generalization of the proximal point algorithm," *SIAM Journal on Control and Optimization*, Vol. 28 (1990), 503-512.

[13] Han, S.P. and Lou, G.: "A parallel algorithm for a class of convex programming," *SIAM*

Journal on Control and Optimization, Vol. 26 (1988), 345-355.
[14] Ibaraki, S., Fukushima, M. and Ibaraki, T.: "Primal-dual proximal point algorithm for linearly constrained convex programming problems," *Computational Optimization and Applications*, Vol. 1 (1992), 207-226.
[15] Ibaraki, S. and Fukushima, M.: "Primal-dual proximal point algorithm for multicommodity network flow problems," *Journal of the Operations Research Society of Japan*, to appear.
[16] Kort, B.W. and Bertsekas, D.P.: "Combined primal-dual and penalty methods for convex programming," *SIAM Journal on Control and Optimization*, Vol. 14 (1976), 268-294.
[17] LeBlanc, L.J., Morlok, E.K. and Pierskalla, W.P.: "An efficient approach to solving the road network equilibrium traffic assignment problem," *Transportation Research*, Vol. 9 (1975), 309-318.
[18] Rockafellar, R.T.: *Convex Analysis*. Princeton University Press, New Jersey, 1970.
[19] Rockafellar, R.T.: "Augmented Lagrangians and applications of the proximal point algorithm in convex programming," *Mathematics of Operations Research*, Vol. 1 (1976), 97-116.
[20] Tseng, P.: "Application of a splitting algorithm to decomposition in convex programming and variational inequalities," *SIAM Journal on Control and Optimization*, Vol. 29 (1991), 119-138.
[21] Zhu, C.: "Modified proximal point algorithm for extended linear-quadratic programming," *Computational Optimization and Applications*, Vol. 1 (1992), 185-206.

Satoru IBARAKI:
Department of Applied Mathematics and Physics,
Faculty of Engineering, Kyoto University,
Kyoto 606, Japan
e-mail: satoru@kuamp.kyoto-u.ac.jp